DOCUMENT RESUME

ED 155 213 TM 007 216

AUTHOR Slaughter, Helen B.; Gallas, Edwin J.
TITLE Will Out-of Level Norm-Referenced Testing Improve the Selection of Program Participants and the Diagnosis of Reading Comprehension in ESEA Title I Programs?
PUB DATE Mar 78
NOTE 26p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978); For related documents, see TM 007 212 and 215

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Achievement Tests; Comparative Testing; Disadvantaged Youth; Equated Scores; Grade 6; *Grade Equivalent Scores; Intermediate Grades; *Low Achievers; Norm Referenced Tests; *Norms; Raw Scores; Reading Achievement; Reading Difficulty; Scores; Standardized Tests; Test Bias; *Testing; *Testing Problems; *Test Interpretation; Test Reliability
IDENTIFIERS California Achievement Tests; Elementary Secondary Education Act Title I; *Out of Level Testing

ABSTRACT

Concern was expressed for the possible effects of testing Elementary Secondary Education Act (ESEA) Title I students with norm-referenced tests that may be so difficult that many students will have scores in the chance range. The likelihood of such students obtaining equal scaled scores if they were tested with easier out-of-level tests was discussed. In this study, two groups of sixth grade students were each tested with two levels of the California Achievement Test. One group took Level 16, on-level, and Level 15, an out-of-level test, while the other group took Level 16 and Level 14. The results are presented in tables. The author concludes that out-of-level testing of low achievers is, at best, a limited solution or stopgap measure to the pressing instrumentation problems in Title I evaluation. (CTM)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

SESSION NO. 31.12

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Helen B. Slaughter
Edwin J. Gallas

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

## Will Out-of-Level Norm-Referenced Testing Improve the Selection of Program Participants and the Diagnosis of Reading Comprehension in ESEA Title I Programs?

Standardized norm-referenced testing programs have generally posed some serious problems for populations that tend to score far below the mean of the national norm group. Test results appear to be less meaningful, and student-teacher frustration increased, as large groups of low-achieving students receive scores that approach the chance level on norm-referenced tests. Test company representatives have suggested and the RMC Research Corporation has recommended that out-of-level testing of students in schools containing large numbers of low achievers would provide a solution to this pressing instrumentation problem, and improve the evaluation of programs designed for low-achieving groups such as ESEA Title I. Both the test companies and the RMC Research Corporation have assumed that tests providing an expanded standard score scale will permit conversions of students' out-of-level raw scores to in-level percentile norms and that the students' percentile rank (or NCE) will be essentially the same as would have been obtained if the appropriate grade-level test had been used (Roberts, 1976).

The major purpose of this study was to test the assumption that groups of students in low-achieving schools would indeed obtain the same interpretive score, e.g., percentile rank and NCE, when testing on reading tests that were designed for students one or more years below their grade placement level as they would have obtained on

grade level tests. After exploring the main effect of out-of-level testing upon mean NCE level, the implications of out-of-level testing for selection of Title I program participants and for improving the assessment of student achievement for very low proficiency readers were determined.

## Method

The 1977 edition of the California Achievement Test, Form C (CAT/C), levels 14, 15 and 16, Reading Comprehension subtest, was used to test sixth grade students in Title I schools. Each student took the on-level test, CAT/C, level 16, designed for grades 5.5 to 6.9 and one of two below-level tests, level 15 designed for grades 4.5-5.9 or level 14 designed for grades 3.5-4.9. The CAT/C was especially designed for functional level testing in that locator tests were provided for selection of the appropriate test level and instructions were written so that multiple levels of the test could be administered concurrently.

A balanced testing design was used in which half of the students in each of the four classes took the on-level test first and the off-level test two days later; the procedure was reversed for the other half of the group. Research Group A (N=47) took the on-level test and the one-level-below-level test (CAT/C 15). Research Group B (N=54) took the on-level test and a test two-levels-below level (CAT/C 14).

Students were given CAT/C Locator Test 1 during the week previous to testing. Locator test scores were not used to assign test levels but were used to explore implications regarding numbers of Title I students who would be assigned lower level tests on this basis,

and to further study the ramifications of individualized testing for Title I populations. (A subsidiary study of individualized testing carried out at three other schools using locator test scores and teacher judgment indicated that teachers invariably based their selection of CAT/C test level upon locator test scores.)

Results and Discussion

Since there were no order effects for either of the research groups, data for students tested either first or second on the in-and-out-of-level tests were combined.

The results shown in Table 1 indicate that low-achieving students tested two-levels-below grade level obtained statistically significant ($p<.001$) lower standard scale and NCE scores than they obtained from on-level tests. Students tested one-level-below grade level also received lower norm-referenced test (NRT) scores on the off-level test than they obtained on the on-level test but the difference was not statistically significant. It is mentioned here because it suggests a systematic trend of decreasing NRT scores accompanying an increase in the number of levels below-grade-level at which students are tested. This may be especially important since Research Group B, taking CAT/C levels 14 and 16, was a lower achieving group than Research Group A and issues of the effects of functional-level testing are especially pertinent to assessing the achievement of that sector of the population. As was expected, students in both research groups achieved significantly higher raw scores on the below-level tests than from the on-level tests indicating that the lower level tests were indeed easier for most students. Correlations between test levels were significantly high.

In interpreting these results, it should be noted that the RMC Research Corporation (Roberts, 1976) recommends out-of-level testing for low-achieving groups whose mean is about a third of a standard deviation higher than the median even if the average raw score of a group is above chance level as were both research groups here. For the combined research groups (N=101), the raw score mean (M=18.35) for CAT/C 16 was approximately one fourth of a standard deviation (SD=8.25) higher than the median (Mdn=16.29). This would suggest that some students were experiencing floor effects but not quite of the magnitude suggested by Roberts for out-of-level testing.

- - - - - - - - - - - - - - -
Insert Table 1 about here
- - - - - - - - - - - - - - -

The RMC technical paper also suggested that it was generally only necessary to drop one test level in out-of-level testing to "improve" the testing situation for low-achieving groups. We examined this presupposition in a number of ways in our study. First we used the CAT/C locator test scores to estimate the number and percentage of Title I students that would have been assigned to lower-level tests in an individualized functional-level testing program. As indicated in Table 2, data regarding test levels prescribed by Locator Test 1 would overwhelmingly select below-level tests for this particular sample of Title I students. The results show that 57 percent would have been tested at two or more levels below grade level and as a whole when the 15 percent prescribed at level 15 were included, 73 percent of the group would have been tested below level.

- - - - - - - - - - - - - - - -
Insert Table 2 about here
- - - - - - - - - - - - - - - -

The question that arises immediately is whether or not such extensive out-of-level testing would have been warranted. For this purpose the data were blocked and analyzed separately for students at three proficiency levels on each level of the test. Students were judged at floor level if they obtained a score that could have been obtained by chance, i.e., 10 or fewer raw score points out of 40 possible points. Our definition of floor level may be low since some measurement experts, e.g., Gulliksen (1950, p. 263), would consider scores within one or even two standard deviations of a chance score to also be at floor level. Students were considered in-range if they obtained a raw score of 11 to 30. Ceiling effects were defined as 31 or more raw score points. Table 3 displays NCE means for proficiency groups defined for both the in-level and out-of-level tests. (Table A in the appendix presents the standard score data.) Eight students in Research Group A and fourteen students in Research Group B, or a total of 22 percent were at the floor level on the CAT/C level 16 test. There is a tremendous difference between 22 percent in need of easier tests based upon empirical results and the 73 percent predicted by the locator test.

- - - - - - - - - - - - - - - - -
Insert Table 3 about here
- - - - - - - - - - - - - - - - -

There has been some speculation that lower norm-referenced scores received by low-achieving groups on off-level tests may more truly reflect the achievement levels of Title I students. We disagree with that proposition. The data presented in Tables 3 and 4 indicate that students who received in-range scores on one or both levels of the test received higher norm-referenced scores, e.g., NCEs or

percentiles, for on-level tests. Students scoring in-range on level 16 received a mean of 5.9 NCEs higher ($p < .001$) for CAT/C level 16 than they received on level 14.

Students scoring in-range and at the floor level on both test levels 14 and 16 also received significantly higher NCE means for the on-level test (Table 4). The other proficiency group means were also higher for the on-level test but differences were not statistically significant. It would seem important in current nationwide focus upon low achievement not to further augment the position of groups who disparage the achievement of today's students.

- - - - - - - - - - - - - - - -
Insert Table 4 about here
- - - - - - - - - - - - - - - -

The number of students at floor level on the off-level test was not remarkably different than that of the on-level test. Nineteen students or three less than for the on-level test also scored at the chance level on the below-level tests! There was even an increase of one floor level score for CAT/C level 15 over level 16. (This can easily have occurred by chance.) Therefore the data presented in Table 3 suggest that testing very low-proficiency readers one or even two test levels below grade placement does not solve the problem of providing adequate instrumentation for testing the reading performance of the 19 to 20 percent of students in the very low-achieving group.

To further explore this question the number of students scoring at the same proficiency level on both in- and off-level tests, especially those at floor level, was examined. Table 4 indicates that 5 students or 11 percent in group A were at floor level on both tests and 7 students or 13 percent in group B were at floor level on both tests. Probably

some degree of test unreliability, often mentioned in connection with low-achievers, combined with exact criterion levels accounts for the reduction from 20 percent of one test score to 12 percent on two test scores being at chance or floor level. (Table B in the appendix presents the standard score data.) As seen in Table 5 a few students in-range on either test scored at the floor on the other regardless of test difficulty.

- - - - - - - - - - - - - - - -
Insert Table 5 about here
- - - - - - - - - - - - - - - -

In order to more clearly understand the phenomenon of "below chance level" achievement test scores, the locator test scores of students scoring at floor level on level 16 or on both levels were examined. All students at floor level on level 16 had received locator test placements indicating a lower level of the CAT/C was needed. All students at floor level on both tests had received locator test placements of at least one level below the lowest level of the test they had taken. Of the five students at floor level on both levels 15 and 16, one received a locator test placement of 12, three were at level 13 and one was at level 14. Of the six of seven students at floor level on levels 14 and 16 for whom locator test scores were available, five placed at level 13 and one at level 12 on the locator test. Of course, of 32 students receiving locator test placements of 12 or 13, only 12 students actually required tests this far removed from grade placement level.

After examining locator test scores, the Title I reading teachers were questioned regarding the characteristics of the floor-level group. Five of the twelve students who scored at the chance

level on both tests were classified as learning disability children. The reading ability of the remaining students was described as low third grade or end of second grade instructional level and their behavior was described as immature or needing a great deal of guidance in the reading process. One child had an unusually limited language and experiential background. Only one child, described as a very good thinker, came close to the category of non-reader. The reading proficiency of students who had been at floor on-level but in-range off-level was judged somewhat higher by the teachers. (Two of these ten children were also learning disability children and one other was making a rapid transition from Spanish to English reading.) In February, teachers tested five children on the level 12 and 13 tests and all but one learning disability student received in-range test scores. It seems that CAT/C, level 13, designed for grade placement 2.5-3.9 may match the students' reading proficiency level according to both teacher judgment and actual test performance.

Table 6 displays data showing the shifts occurring between percentages of students scoring at various proficiency levels with out-of-level testing. It appears that there is a greater shift towards more ceiling effects with out-of-level testing than there is a decrease in the percentage of floor effects. Further, only 69 percent of the group scored at the same proficiency level on both tests.

- - - - - - - - - - - - - - - -
Insert Table 6 about here
- - - - - - - - - - - - - - - -

In terms of stanine, NCE and percentile rank equivalents of raw score intervals for on- and off-level tests, Table 7 presents data indicating that a greater probability, i.e., greater stanine range, of

ceiling than floor effects can be expected for all levels on this particular subtest. For example, chance raw scores extend into stanine two for level 16 while ceiling effects include stanines seven through nine. (By the end of grade six, ceiling effects extend into stanine six.) Chance raw scores for level 14, two levels below grade placement, are confined within stanine one while ceiling effects include stanines five through nine. This shows the magnitude of raw score points required to obtain a similar percentile rank on a lower level test than would be required on an on-level test.

- - - - - - - - - - - - - - - -
Insert Table 7 about here
- - - - - - - - - - - - - - - -

Students are sometimes selected for participation in Title I programs on the basis of standardized test scores. A stanine range of 1-3 is a common criterion for defining the Title I target population. The data present in Table 8 show that if students are tested one level below grade placement there will be an increase of six percent in the target population. The increase for Research Group A was from 28 percent for on-level testing to 34 percent for off-level testing. In Research Group B, 57 percent of students tested two levels below grade placement would be selected as target students while only 46 percent were selected when tested on-level. Therefore, testing two levels below resulted in increasing the target group by 11 percent.

- - - - - - - - - - - - - - - -
Insert Table 8 about here
- - - - - - - - - - - - - - - -

The combined research groups increased the target population by 9 percent when testing one or two levels below level, an increase from 38 percent on-level to 47 percent off-level. Although groups A and B

were not equal. they are both part of the eligible target school population in one district. If an individualized testing program were implemented rather than a controlled design, a target population selected upon the basis of test scores would be increased relative to the numbers of lower level tests administered.

Conclusions and Recommendations

Some previous research comparing the out-of-level test scores of Title I students in Rhode Island with on-level test scores found that different grade-equivalent scores were obtained for off-level and on-level tests (Long, Schaffran and Kellogg, 1977). Grade-equivalent scores were lower for grade four students when tested below-level and higher for grades two and three students when tested below-level. The findings of this study agree with the Rhode Island fourth grade results in that sixth grade students in the Tucson Unified School District Title I program obtained lower derived scores, i.e., NCEs, when tested below level than when tested at grade placement level. While that study used the Gates-MacGinitie Reading Test and this one used the CAT/C Reading Comprehension subtest both studies suggest a need to reexamine the procedures used to develop norms for tests to be used at both grade and functional level. Rather than relying solely upon overlapping each level of the CAT/C in the spring with the next higher test level in constructing the expanded standard score scales, empirical norms should be developed for out-of-level testing by testing students at multiple test levels in the norming procedure. If a systematic bias to derived scores of out-of-level tests remains a problem, perhaps a regression formula could be used to provide an estimated in-level

derived score. As stated by Long et al., (1977, p. 212), "Clearly this study indicates weaknesses in cross-form scaling of the test used."

The issue of test fairness to low-achieving students is becoming increasingly important as more political pressure is put upon educators to maintain minimum standards in basic skill areas. As was found in this study and in the Rhode Island study, even larger percentages of students would be selected for remedial-type programs on the basis of off-level tests than on the basis of grade-level tests. It would seem that the last thing a low achiever needs is a test that is even more difficult to "pass" than those currently in use.

A reasonable plan for Title I testing would seem to be to test most students on-level and to retest only students scoring below chance level on locator tests and the prescribed lower-level test. However, the data from on- and off-level tests should be kept separate for both individual and group evaluation purposes. If a test were administered one level below level, the use of a percentile band for individual scores would probably suffice; however the actual need seems to be for tests two or more levels below grade placement for the very low proficiency reader. (The CAT/C has more refined levels than other test batteries so the number of levels below grade placement used will be dependent upon the test battery.) Also, the effects of out-of-level testing will have to be researched for each different grade level and test battery.

The value of out-of-level testing is that it is one way to measure the achievement of some very low proficiency readers rather than to treat all low achievement as a void. It is important to obtain

in-range scores for as large a number of students as possible. Yet, we must question the adequacy of the lower levels of the CAT/C as a criterion-referenced measure of the reading proficiency of older low achievers. The term functional level testing has been avoided in connection with this study as the CAT/C and other survey tests may not be comprehensive enough to adequately measure functional reading proficiency, especially for students at increasingly higher grade levels. In fact, there is probably a trade-off between how well a lower-level test serves as an appropriate measure of functional reading performance and the congruence of derived off- and on-level test scores. Reading is measured in a somewhat different way in the CAT/C level 12 and 13 tests than in level 16. At the lower levels there are fewer comprehension items following passages and sentence length comprehension, structural analysis and word recognition items are included. The tests also require a guided testing technique where oral directions are frequently given during the testing sessions. Therefore the levels 12 and 13 approach to testing seems appropriate to reading behaviors of "chance level" students as described by the reading teachers.

It appears that out-of-level testing of low-achievers is at best a limited solution or stopgap measure to the pressing instrumentation problems in Title I evaluation. One is reminded of the dilemma of a chef who has a saucepan without a handle which s/he will not throw away for lack of a replacement. Yet the saucepan is very awkward to use. One can use it but it is tricky. That is the situation at present for anyone trying to interpret test results from out-of-level tests.

## REFERENCES

California Achievement Tests, Form C, CTB/McGraw-Hill, Del Monte Research Park, Monterey, California, 1977.

Gulliksen, H. Theory of Mental Tests. New York: John Wiley & Sons, Inc., 1950.

Long, J. V., Schaffran, J. A., & Kellogg, T. M. Effects of out-of-level survey testing on reading scores of Title I, ESEA students. Journal of Educational Measurement. (14) (3), fall 1977, pp. 203-213.

Roberts, A. O. H. Out of-level testing. ESEA Title I Evaluation and Reporting System. Technical Paper no. 6. Mountain View, California: RMC Research Corporation. October, 1976.

Table 1. Comparison of Reading Achievement Scores of Sixth Grade Students Tested On and Below Grade Level on the CAT/C[1]

| | N | RS M | RS SD | SS M | SS SD | NCE M | NCE SD | Percentile Equivalents of NCE Means |
|---|---|---|---|---|---|---|---|---|
| Group A | | | | | | | | |
| Level 15 | 47 | 22.3 | 10.1 | 475.2 | 71.9 | 44.2 | 22.0 | 38.4 |
| Level 16 | 47 | 20.7 | 8.6 | 481.3 | 63.0 | 46.2 | 20.6 | 43.2 |
| Difference | | -1.6* | | 6.1 | | 2.0 | | 4.8 |
| Group B | | | | | | | | |
| Level 14 | 54 | 20.0 | 9.1 | 435.1 | 58.4 | 31.1 | 17.9 | 18.5 |
| Level 16 | 54 | 16.3 | 7.4 | 452.3 | 55.2 | 36.4 | 17.2 | 25.9 |
| Difference | | -3.7*** | | 17.2*** | | 5.3*** | | 7.4 |
| Correlations | | | | | | | | |
| Levels 15 and 16 | | .88*** | | .89*** | | .84*** | | |
| Levels 14 and 16 | | .86*** | | .83*** | | .83*** | | |

[1]Note: Correlated *t* tests were used to compare the significance of the difference between means for all but the percentile equivalents of NCEs.

* $p < .05$

*** $p < .001$

Table 2. CAT/C Test Levels Prescribed by Locator Test 1

| Group | CAT/C Levels 12 N | % | 13 N | % | 14 N | % | 15 N | % | 16 N | % | 17 N | % | Total N* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 2 | 5 | 9 | 21 | 11 | 25 | 8 | 19 | 10 | 23 | 3 | 7 | 43 |
| B | 3 | 6 | 18 | 37 | 9 | 19 | 6 | 12 | 11 | 24 | 1 | 2 | 48 |
| Total | 5 | 6 | 27 | 30 | 20 | 22 | 14 | 15 | 21 | 23 | 4 | 4 | 91 |

* Some students were absent on the day the locator test was administered

Table 3. Comparison of NCE Means for Groups Blocked on Floor, Ceiling and In-Range Proficiency Levels and In-and-Out-of-Level Tests

| Test | Floor N | Floor M | In-Range N | In-Range M | Ceiling N | Ceiling M | Total N | Total M |
|---|---|---|---|---|---|---|---|---|
| **Group A Proficiency Groups Via CAT/C Level 16** | | | | | | | | |
| CAT/C 15 | 8 | 17.6 | 32 | 44.1 | 7 | 75.0 | 47 | 44.2 |
| CAT/C 16 | 8 | 15.6 | 32 | 47.1 | 7 | 77.4 | 47 | 46.2 |
| Difference | | -2.0 | | 3.0 | | 2.4 | | 2.0 |
| **Group A Proficiency Groups Via CAT/C Level 15** | | | | | | | | |
| CAT/C 15 | 9 | 14.3 | 25 | 40.5 | 13 | 71.9 | 47 | 44.2 |
| CAT/C 16 | 9 | 22.3 | 25 | 43.5 | 13 | 68.2 | 47 | 46.2 |
| Difference | | 8.0 | | 3.0 | | 3.7 | | 2.0 |
| **Group B Proficiency Groups Via CAT/C Level 16** | | | | | | | | |
| CAT/C 14 | 14 | 14.2 | 37 | 34.9 | 3 | 63.9 | 54 | 31.1 |
| CAT/C 16 | 14 | 16.4 | 37 | 40.8 | 3 | 75.6 | 54 | 36.4 |
| Difference | | 2.2 | | 5.9*** | | 11.7 | | 5.3*** |
| **Group B Proficiency Groups Via CAT/C Level 14** | | | | | | | | |
| CAT/C 14 | 10 | 6.2 | 35 | 31.2 | 9 | 58.8 | 54 | 31.1 |
| CAT/C 16 | 10 | 21.5 | 35 | 34.2 | 9 | 61.3 | 54 | 36.4 |
| Difference | | 15.3** | | 3.0 | | 2.5 | | 5.3*** |

** $p < .01$

*** $p < .001$

Table 4. Comparison of NCE Means for Groups of Grade Six Students Scoring at the Same Proficiency Levels on CAT/C Levels 14, 15, and 16

| Group A (N=34) Proficiency Groups | One-Level-Below CAT/C 15 | | | On-Level CAT/C 16 | | |
|---|---|---|---|---|---|---|
| | N | M | SD | M | SD | Difference |
| Floor | 5 | 11.9 | 7.6 | 13.6 | 8.0 | 1.7 |
| In-Range | 22 | 42.3 | 9.2 | 46.8 | 12.1 | 4.5 |
| Ceiling | 7 | 75.0 | 10.7 | 77.4 | 5.1 | 2.4 |

| Group B (N=36) Proficiency Groups | Two-Levels-Below CAT/C 14 | | | On-Level CAT/C 16 | | |
|---|---|---|---|---|---|---|
| | N | M | SD | M | SD | Difference |
| Floor | 7 | 6.2 | 5.2 | 16.5 | 5.1 | 10.3* |
| In-Range | 27 | 32.9 | 7.9 | 37.5 | 7.5 | 4.6*** |
| Ceiling | 2 | 72.5 | 17.1 | 78.3 | 16.0 | 5.8 |

* $p < .05$

*** $p < .001$

Table 5. Comparison of NCE Means for Students at Floor and In-Range on In-and-Off Level Tests

| Group A | N | CAT/15 M | CAT/15 SD | CAT/16 M | CAT/16 SD | Difference |
|---|---|---|---|---|---|---|
| Floor on 16, In-Range on 15 | 3 | 27.1 | 3.3 | 18.9 | 5.0 | -8.2 |
| Floor on 15, In-Range on 16 | 4 | 17.3 | --- | 33.2 | 5.4 | 15.9 |

| Group B | N | CAT/14 M | CAT/14 SD | CAT/16 m | CAT/16 SD | Difference |
|---|---|---|---|---|---|---|
| Floor on 16, In-Range on 14 | 7 | 22.1 | 5.2 | 16.4 | 7.9 | -5.7 |
| Floor on 14, In-Range on 16 | 3 | 6.1 | 4.8 | 33.2 | 2.9 | 27.1 |

Table 6. Number and Percentage of 101 Students at Proficiency Levels on In-Level, Out-of-Level and Both Levels of the CAT/C

| | Both Levels | | In-Level | | Out-of-Level | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| In-Range | 49 | 49 | 69 | 68 | 60 | 59 |
| Ceiling | 9 | 9 | 10 | 10 | 22 | 22 |
| Floor | 12 | 12 | 22 | 22 | 19 | 19 |
| Total | 70 | 69 | 101 | 100 | 101 | 100 |

Table 7. Raw Score Cutoff Points for Stanines, Percentile Rank and Nearest Whole NCE for Grade 6.1 Students for CAT/C, Levels 14, 15 and 16, Reading Comprehension Subtest[1].

| | Stanine, Percentile Rank and NCE | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Stanine | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Percentile Rank | 1- 4 | 5-10 | 11-22 | 23-40 | 41-59 | 60-77 | 78-89 | 90-95 | 96-99 |
| NCE | 1-13 | 15-23 | 24-34 | 34-45 | 45-55 | 55-66 | 66-76 | 77-85 | 87-99 |
| CAT/C 16 RS | 0- 8* | 9-10* | 11-14 | 15-20 | 21-25 | 26-30 | 31-34** | 35-36** | 37-40** |
| CAT/C 15 RS | 0- 9* | 10-11* | 12-16 | 17-23 | 24-29 | 30-34** | 35-36** | 37-38** | 39-40** |
| CAT/C 14 RS | 0-10* | 11-14 | 15-21 | 22-28 | 29-33** | 34-36** | 37-38** | 39** | 40** |

[1]Note: Off-level raw scores were converted to off-level standard scores which were converted to on-level percentiles for grade 6.1.

*Chance level raw score range 1-10.

**Ceiling level raw score range 31-40.

Table 8. Size of Target Groups, Stanines 1-3, for Students Tested at Nominal level and One or Two levels Below Nominal Level

| Research Group A | N | % |
|---|---|---|
| Total Group tested | 47 | |
| On Level (CAT/C, 16) target group | 13 | 28 |
| One Level Below (CAT/C, 15) target group | 16 | 34 |
| Difference | | 6 |
| **Research Group B** | | |
| Total Group tested | 54 | |
| On Level (CAT/C, 16) target group | 25 | 46 |
| Two Levels below (CAT/C, 14) target group | 31 | 57 |
| Difference | | 11 |
| **Combined Research Groups** | | |
| Total Group tested | 101 | |
| On Level (CAT/C, 16) target group | 38 | 38 |
| Off Level (CAT/C, 15 & 14) target group | 47 | 47 |
| Difference | | 9 |

# APPENDIX

Table A. Comparison of Standard Score Means for Groups Blocked on Floor, Ceiling and In-Range Proficiency Levels on In-and-Out-of-Level Tests.

| Group A Tests | Floor | | In-Range | | Ceiling | | Total | |
|---|---|---|---|---|---|---|---|---|
| | N | M | N | M | N | M | N | M |
| | | | Proficiency Groups Via Level 16 Test | | | | | |
| CAT/C 15 | 8 | 387.9 | 32 | 474.9 | 7 | 576.0 | 47 | 475.2 |
| CAT/C 16 | 8 | 387.5 | 32 | 482.4 | 7 | 584.1 | 47 | 481.4 |
| Diff. | | -.4 | | 7.5 | | 8.1 | | 6.2 |
| | | | Proficiency Groups Via Level 15 Test | | | | | |
| CAT/C 15 | 9 | 377.3 | 25 | 463.0 | 13 | 565.9 | 47 | 475.2 |
| CAT/C 16 | 9 | 408.2 | 25 | 470.1 | 13 | 553.8 | 47 | 481.4 |
| Diff. | | 30.4 | | 7.1 | | -12.1 | | 6.2 |

| Group B Tests | Floor | | In-Range | | Ceiling | | Total | |
|---|---|---|---|---|---|---|---|---|
| | N | M | N | M | N | M | N | M |
| | | | Proficiency Groups Via Level 16 Test | | | | | |
| CAT/C 14 | 14 | 379.6 | 37 | 447.7 | 3 | 539.7 | 54 | 435.1 |
| CAT/C 16 | 14 | 387.5 | 37 | 466.7 | 3 | 577.7 | 54 | 452.3 |
| Diff. | | 7.9 | | 19.0 | | 38.0 | | 17.2 |
| | | | Proficiency Groups Via Level 14 Test | | | | | |
| CAT/C 14 | 10 | 352.3 | 35 | 435.7 | 9 | 524.9 | 54 | 435.1 |
| CAT/C 16 | 10 | 403.6 | 35 | 445.8 | 9 | 531.8 | 54 | 452.3 |
| Diff. | | 51.3 | | 10.1 | | 6.9 | | 17.2 |

Table B. Comparison of Standard Scores for Proficiency Groups of Students on Two Test Levels.

| | Below Level | | | CAT/16 | | Diff. |
|---|---|---|---|---|---|---|
| Group A (15/16) Proficiency Groups | N | M | SD | M | SD | |
| Floor, Both Levels | 5 | 367.2 | 34.9 | 381.2 | 23.1 | 14.0 |
| In-Range, Both Levels | 22 | 468.0 | 29.0 | 480.0 | 28.5 | 11.4 |
| Ceiling, Both Levels | 7 | 576.0 | 35.5 | 584.1 | 17.0 | 8.1 |
| Floor on 16, In-Range on 15 | 3 | 422.3 | 11.7 | 398.0 | 15.6 | -24.3 |
| Floor on 15, In-Range on 16 | 4 | 391.0 | | 442.0 | 17.3 | 51.0 |
| Group B (14/16) Proficiency Groups | | | | | | |
| Floor, Both Levels | 7 | 352.1 | 16.8 | 386.8 | 18.7 | 34.7 |
| In-Range, Both Levels | 27 | 441.3 | 24.9 | 456.6 | 23.4 | 15.3 |
| Ceiling, Both Levels | 2 | 566.5 | 54.4 | 587.5 | 54.4 | 21.0 |
| Floor on 16, In-Range on 14 | 7 | 407.0 | 16.8 | 388.1 | 24.6 | -18.9 |
| Floor on 14, In-Range on 16 | 3 | 352.7 | 11.5 | 442.7 | 8.1 | 90.0 |